

23.11.2017 – [DHnet|Jena](#)
Thementag Digital Humanities in Jena

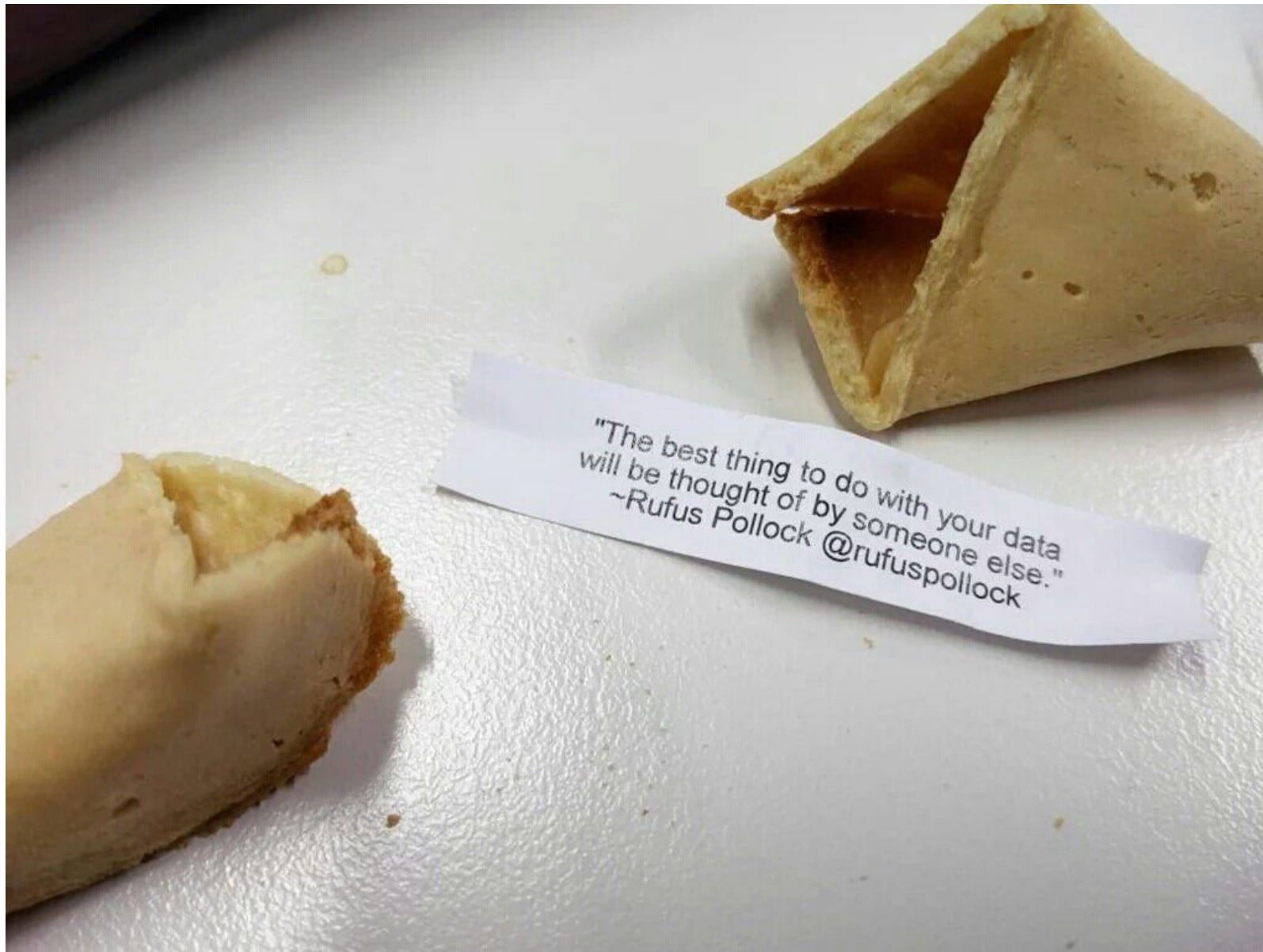
Set your data free!

VOM GEISTESWISSENSCHAFTLICHEN
DATENSILO ZUR VERNETZTEN
WISSENSBASIS MITTELS LINKED OPEN
DATA

Torsten Schrade (<https://orcid.org/0000-0002-0953-2818>)

Akademie & Hochschule Mainz  [@digicademy](#) & [@_mainzed](#) | [CC-BY 4.0](#)

MOTTO



Quelle: [Rufus Pollock](#), Gründer der [Open Knowledge Foundation](#)

GLIEDERUNG

1. **Einführung**

Vom Babylon der Forschungsdaten

2. **Modellierung**

Die Welt (und Forschungsdaten) in Begriffe fassen

3. **Praxis**

Anwendungsbeispiele

4. **Zusammenfassung**

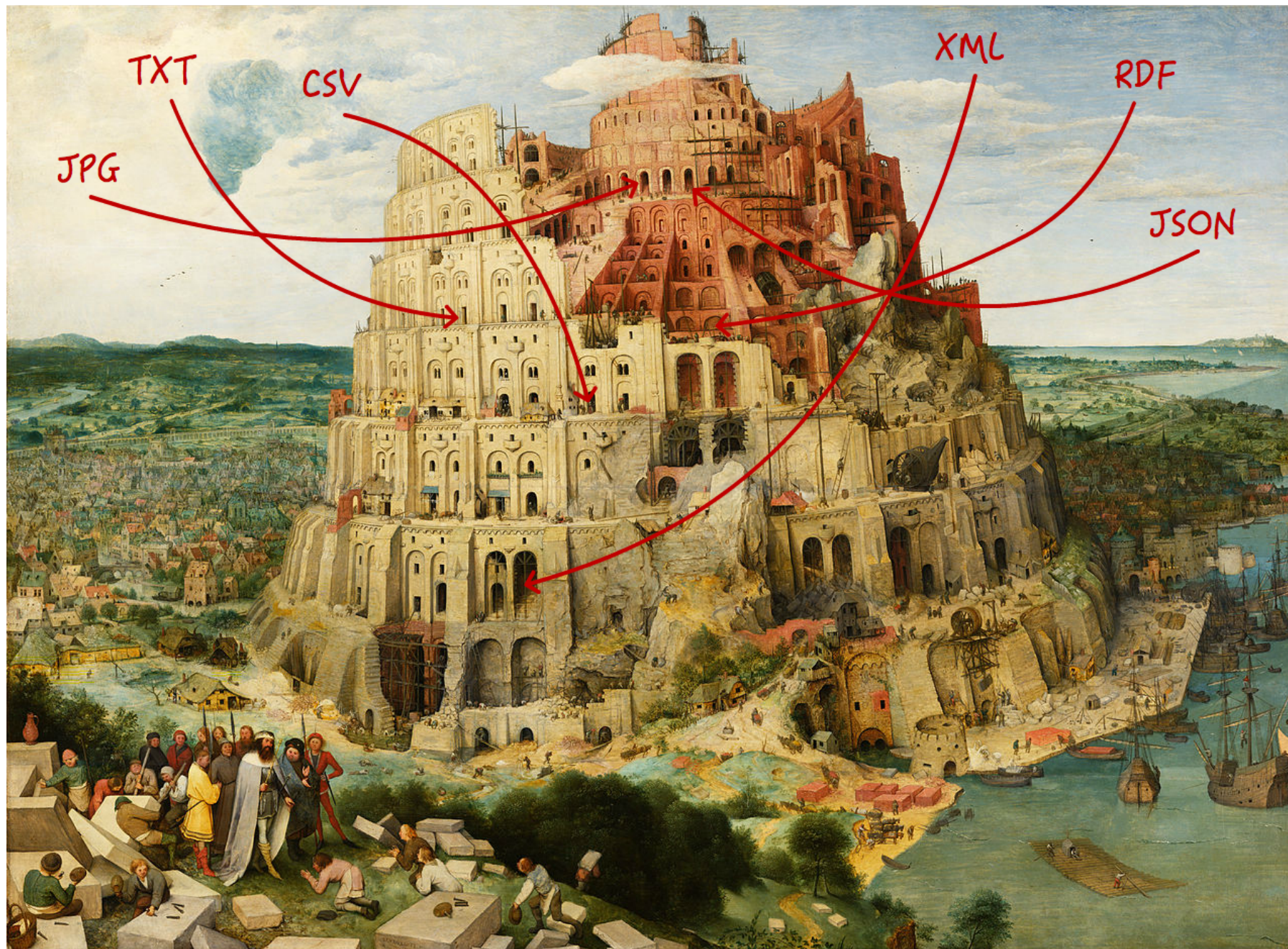
Potentiale von LOD in den Digital Humanities

01 Einführung

VOM BABYLON DER
FORSCHUNGSDATEN

HABEN SIE AUCH FORSCHUNGSDATEN?

Heterogene Datenquellen und -formate



TEI ≠ TEI

Heterogene Modellierungsweise gleicher Phänomene

The image displays three browser windows side-by-side, each showing a different TEI XML encoding for the same archaeological object. The windows are titled 'www.inschriften.net/rest/di060/articles/di060-', 'file:///Users/schrade/Desktop/mz2-1023.xr', and 'inslib.kcl.ac.uk/irt2009/IRT001:'. Each window shows a different XML structure, with red boxes highlighting specific sections that illustrate heterogeneity in modeling.

Window 1 (Left): Shows a TEI encoding with a `<history>` section highlighted in red. The `<history>` section contains `<origin>` and `<origDate>` elements. The `<origDate>` element has a `date` attribute with a range from "1606-01-01" to "1606-12-31" and a `datingPoint` attribute with a value of "1606AAA0000000A3".

Window 2 (Middle): Shows a TEI encoding with a `<history>` section highlighted in red. The `<history>` section contains `<origin>` and `<origDate>` elements. The `<origDate>` element has a `date` attribute with a value of "1094/95" and a `notBefore` attribute with a value of "1094/95".

Window 3 (Right): Shows a TEI encoding with a `<div type="description" subtype="date">` section highlighted in red. The `<div>` section contains a `<date type="textDate" notBefore="0101" notAfter="0188" exact="none">` element, which is followed by a `<rs type="criteria">` element with a value of "lettering".

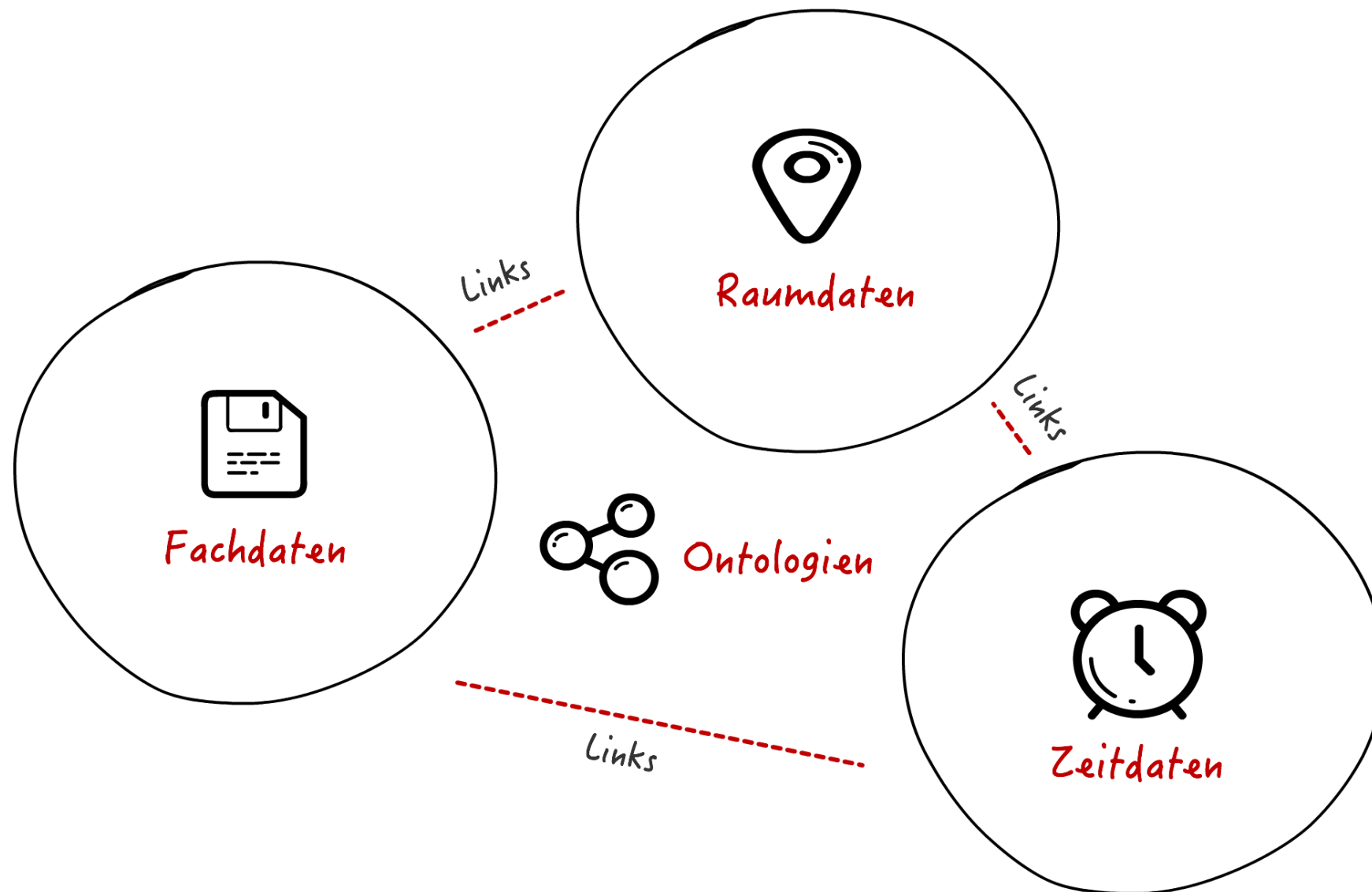
RESULTAT: DATENSILOS

Eine "lonely crowd" an Forschungsdaten



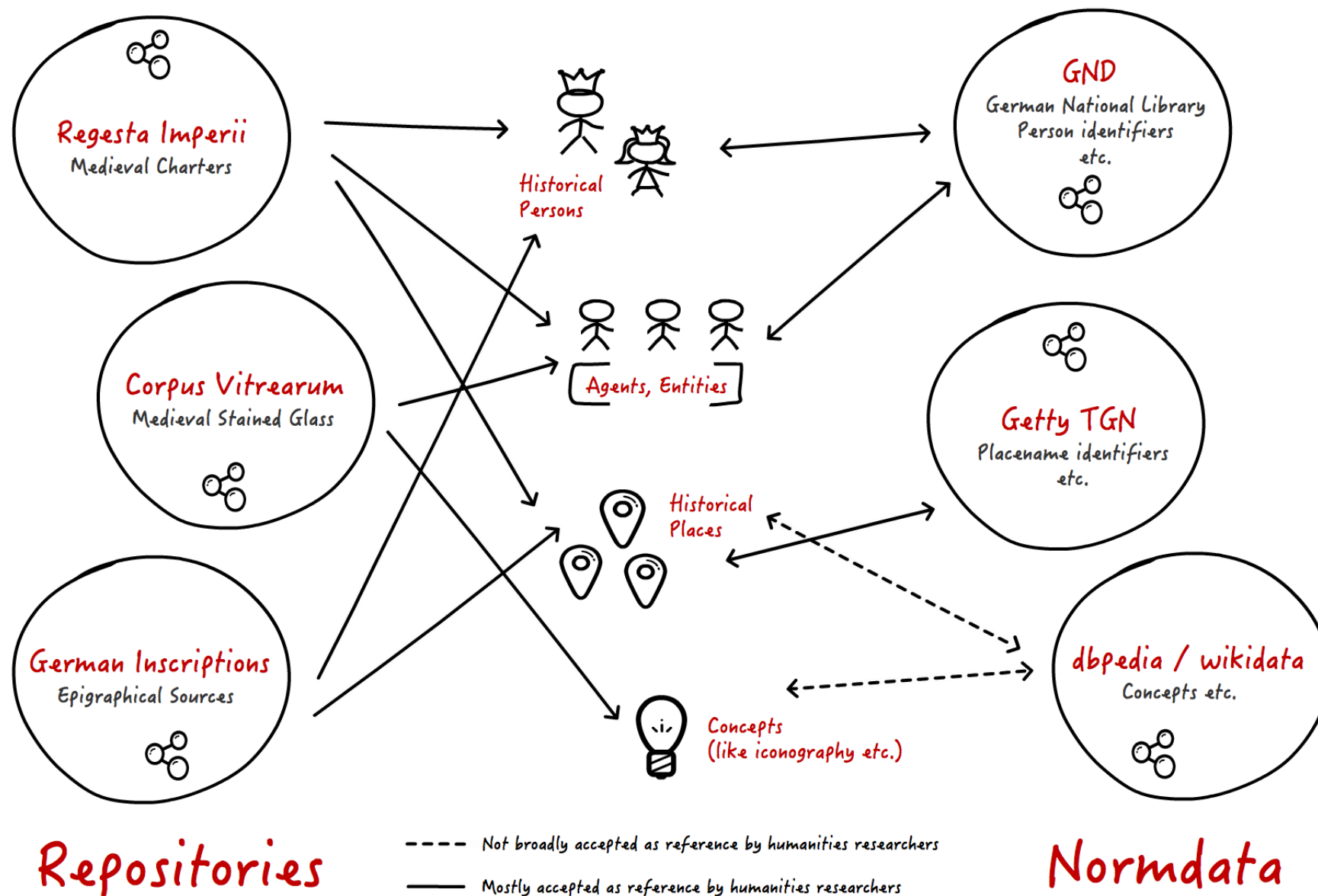
SEMANTISCHE VERKNÜPFUNG

Übergreifende Analyse von Forschungsdaten



BEGRIFFE UND KONZEPTE

Verknüpfung mittels Normdaten, Vokabularen & Ontologien



KATEGORIEN

Grundlage einer Wissensrepräsentation

*Aristoteles (384-322 BCE): Kategorien als "**Aussageschemata** in Bezug auf etwas **Seiendes**. Aristoteles führt mit den Kategorien eine neue Art logischer Ausdrücke ein, die es erlaubt **Prädikate** von **Subjekten** zu unterscheiden und die Prädikate logisch zu klassifizieren. Es geht ihm darum, grundlegend und vollständig die **Formen, nach denen einzelne Wörter sinnvoll zu Aussagen verbunden werden können**, zu bestimmen. Seine Einteilung der Aussageschemata orientiert sich dabei an einfachen, nicht-zusammengesetzten sprachlichen Ausdrücken wie „Sokrates“, „Mensch“ oder „weiß“. Da Ausdrücke, Begriffe und Bezeichnetes nicht immer klar unterschieden sind, ist auch eine Lesart möglich, nach der **zugleich die Wirklichkeit in Typen von Objekten und Eigenschaften** eingeteilt werden (im Sinne einer Ontologie)." ([Wikipedia](#))*

KATEGORIEN

Grundlage einer Wissensrepräsentation

- ▶ Ding (substantia)
- ▶ Größe (quantitas)
- ▶ Beschaffenheit (qualitas)
- ▶ Beziehung (relatio)
- ▶ Ort (ubi)
- ▶ Zeit (quando)
- ▶ Lage (situs)
- ▶ Haben (habitus)

- ▶ Tun (actio)
- ▶ Erleiden (passio)

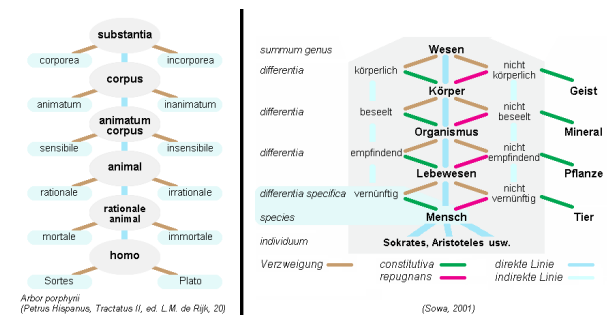
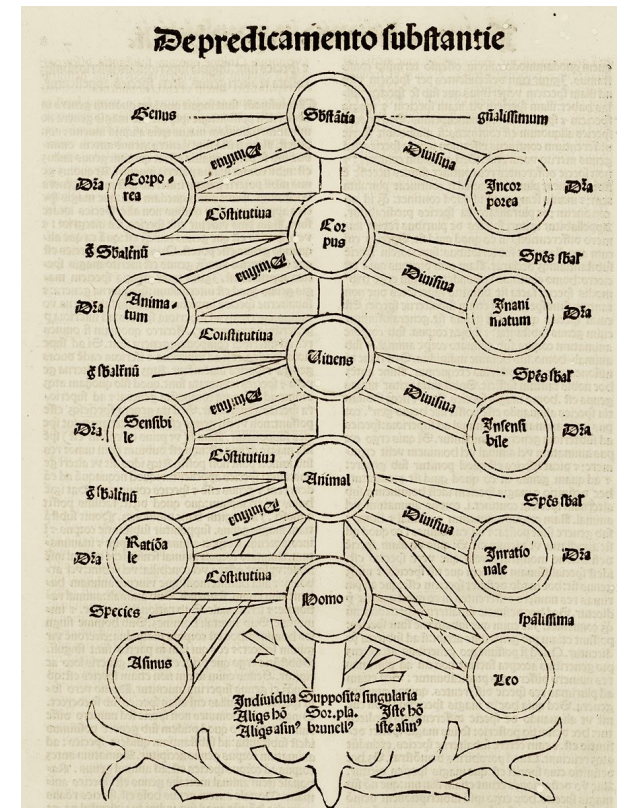
KATEGORIEN

Baum des Wissens als epistemologisches Ordnungssystem

Auf Grundlage der Isagoge des Porphyrius (~270 CE): Hierarchie von Arten/Gattungen (Holzschnitt aus: Boëthius, Opera varia. Pars I. Venedig 1497).

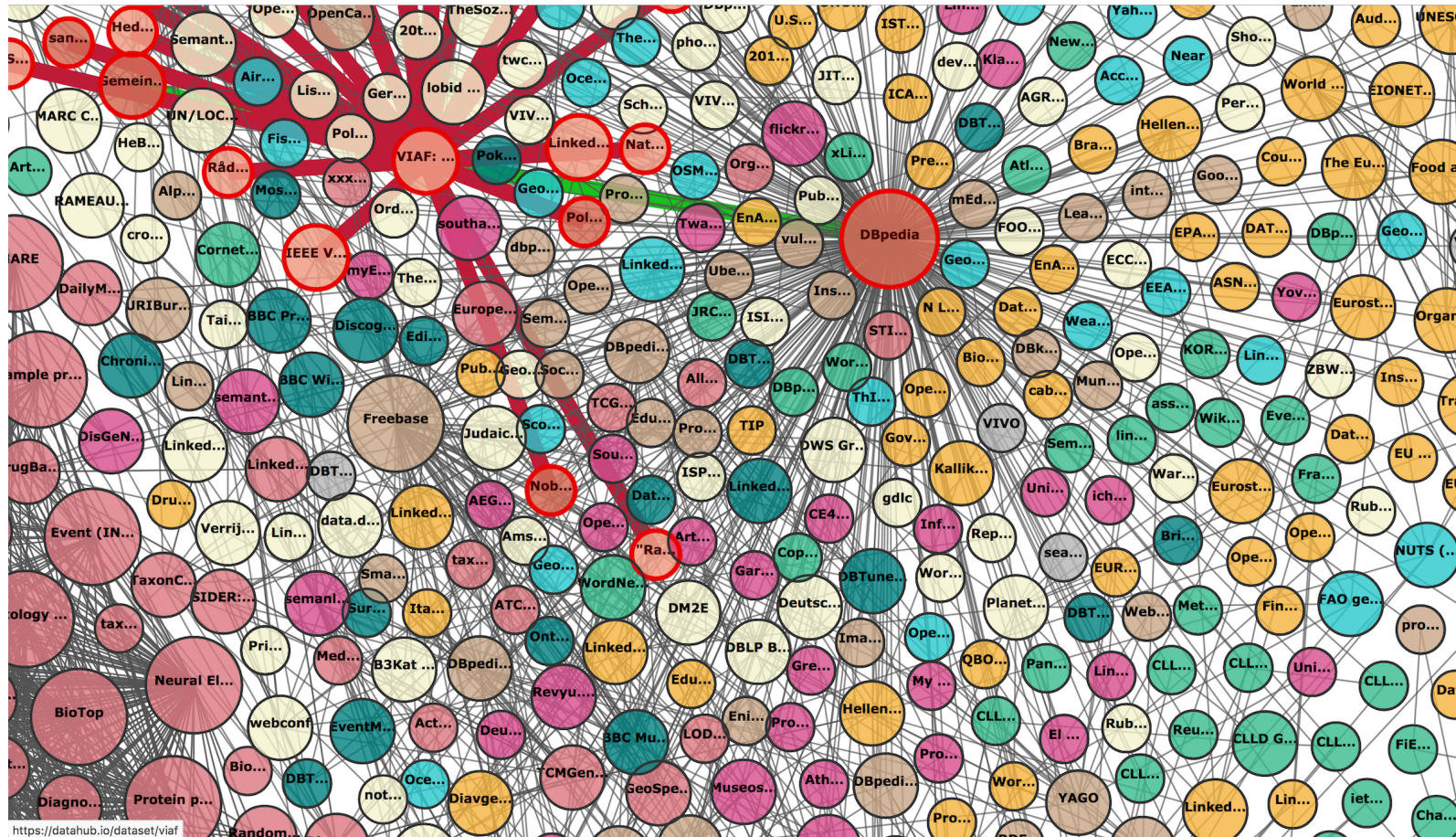
„Die jeweils höchste Gattung (das *summum genus*) eines solchen Baumes ist die Kategorie. Sie bestimmt den höchsten Abstraktionsgrad. Im Gegensatz zu darunter liegenden Ebenen kann die höchste Gattung nicht Art einer anderen sein. Eine niedrigste Art (*infima species*)

kann im Gegensatz zu darüber liegenden Ebenen nicht mehr weiter eingeteilt werden. Es handelt sich um einen Individualbegriff.“ ([Wikipedia](https://de.wikipedia.org/wiki/Arbor_porphyrii))



DAS SEMANTIC WEB

Ein „Giant Global Graph“



Quelle: Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>

02

MODELLIERUNG

Die Welt (und Forschungsdaten) in Begriffe fassen

GRUNDLAGEN

Vom „Web of Documents“ zum „Web of Data“

- ▶ Vorschlag von Tim Berners-Lee: "The Semantic Web is an **extension of the current web** in which information is given well-defined meaning, better enabling computers and people to work in cooperation" ([Quelle](#))
- ▶ Das "Five-Star" LOD Prinzip von Berners-Lee: ([Quelle](#))
 - ▶ ★ Stelle Daten im Web unter einer offenen Lizenz bereit. Das Format ist dabei egal
 - ▶ ★★ Stelle Daten in einem strukturierten Format bereit
 - ▶ ★★★ Verwende offene, nicht proprietäre Formate
 - ▶ ★★★★ Verwende URIs um Dinge zu bezeichnen
 - ▶ ★★★★★ Verlinke deine Daten mit anderen Daten um Kontexte herzustellen

GRUNDLAGEN

Technische Hintergründe

- ▶ Das Semantic Web baut auf den **existierenden Web-Standards** (URIs, HTTP, HTML, u.a.) auf
- ▶ Durch ein **standardisiertes Datenmodell (RDF)** und eine **standardisierte Ontologie (OWL)** wird es möglich, implizite Informationen in und über (Web)Ressourcen zu explizieren, übergreifend miteinander zu verbinden und auch für eine maschinelle Auswertung verfügbar zu machen
- ▶ **RDF ist ein Datenmodell.** Es legt nicht fest, in welcher Form es konkret *serialisiert* wird. Bis heute haben sich zahlreiche **Serialisierungsformate** für RDF entwickelt (RDF/XML, Turtle, JSON-LD etc.)
- ▶ Die Grundeinheit des Semantic Web bildet das **Triple**. Ein Triple stellt eine **Aussage** bestehend aus einem **Subjekt**, einem **Prädikat** und einem **Objekt** dar.

GRUNDBEGRIFFE

Triple (Prinzip)

<Goethe>

<a>
<sent>

<Sender> ;
<Letter> .

<Letter>

<dateSent>
<sentFrom>

"1814" ;
<Wiesbaden> .

<Wiesbaden>

<a>
<lat>
<long>

<SpatialThing> ;
"50.08" ;
"8.24" .

■ Subjekt ■ Prädikat ■ Objekt <...> = URIs, "... " = Literale

GRUNDBEGRIFFE

Triple (RDF)

```
<http://d-nb.info/gnd/118540238> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <https://lod.academy/cmi/vocab/#S  
    <https://lod.academy/cmi/vocab/#sent> <http://www.weber-gesamtausgabe.c  
  
<http://www.weber-gesamtausgabe.de/A040654> <https://lod.academy/cmi/vocab/#dateSent> "1814" ;  
    <https://lod.academy/cmi/vocab/#sentFrom> <http://www.geonames.org/28093  
  
<http://www.geonames.org/2809346> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <https://www.w3.org/2003/01/geo/  
    <https://www.w3.org/2003/01/geo/wgs84_pos#lat> "50.08" ;  
    <https://www.w3.org/2003/01/geo/wgs84_pos#long> "8.24" .
```

■ Subjekt ■ Prädikat ■ Objekt

GRUNDBEGRIFFE

URI

The diagram shows the URI `foo://example.com:8042/over/there?name=duck#beak` with its components labeled below it. The labels are: `foo` is the `scheme`; `://example.com:8042` is the `authority`; `/over/there` is the `path`; `?name=duck` is the `query`; and `#beak` is the `fragment`. Brackets and lines connect each label to its corresponding part of the URI.

Ein URI besteht nach [RFC 3986](#) aus fünf Teilen: `scheme` (Schema oder Protokoll), `authority` (Anbieter oder Server), `path` (Pfad), `query` (Abfrage) und `fragment` (Teil). Nur **scheme, authority und path** müssen in jedem URI vorkommen.

Berners-Lee: *Cool URIs don't change!*

GRUNDBEGRIFFE

Vokabulare und Ontologien

- ▶ Ontologien in der Informatik modellieren **Begriffe bzw. Informationskonzepte** und deren **jeweilige Beziehungen** untereinander
- ▶ **Geisteswissenschaftliche Informationskonzepte** können bspw. sein:
 - ▶ Personen und deren Familienbeziehungen untereinander
 - ▶ Ereignisse und deren Auswirkungen
 - ▶ Kunstwerke bzw. Artefakte, deren Entstehung und Eigenschaften
 - ▶ ... etc.
- ▶ **Hilfreiche Ontologien** für die Geistes- und Kulturwissenschaften sind u.a.:
 - ▶ **FOAF** (Personen)
 - ▶ **REL** (Beziehungen zwischen Personen)
 - ▶ **BIO** (Eigenschaften von Personen)
 - ▶ **CIDOC-CRM** bzw. **Erlangen-CRM** (Kulturelles Erbe)
 - ▶ ... und viele viele Weitere. **Ontologien finden:** <http://prefix.cc/>

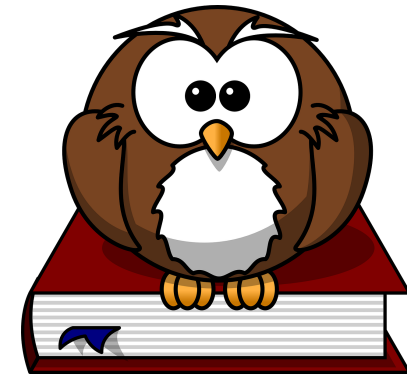


Vokabulare und Ontologien
modellieren Konzepte

GRUNDBEGRIFFE

RDFS & OWL

- ▶ RDF legt nur ein gemeinsames Modell bzw. eine Syntax für den Datenaustausch fest. Zur (maschinellen) Interpretation bzw. **formalen Beschreibung** der in RDF getroffenen Aussagen werden Vokabulare benötigt. Nur dadurch wird eine maschinelle Auswertbarkeit möglich.
- ▶ RDF-Schema (RDFS) stellt ein Vokabular zur Verfügung, mit dem die in einer Fachdomäne vorkommenden **Begriffe und deren Eigenschaften und Relationen** modelliert werden können. Mit RDFS lassen sich einfache Ontologien realisieren.
- ▶ Enthält ein Vokabular gleichzeitig auch **Regeln für die richtige Verwendung** der in ihm definierten Begriffe spricht man von einer Ontologie.
- ▶ Komplexe Ontologien benötigen eine formale Beschreibungssprache, die ausdrucksmächtiger ist als RDF-Schema. Hierfür existiert die **Web Ontology Language (OWL)** des W3C.
- ▶ Durch die Verwendung von OWL innerhalb einer Ontologie wird es möglich, **logische Schlussfolgerungen** auf den Daten auszuführen, die mit dieser Ontologie arbeiten. Hierbei gilt die **Open World Assumption**: *Ein "Reasoner" nimmt an, dass etwas existiert, solange nicht explizit definiert wurde, dass es nicht existiert.*



OWL (Web Ontology Language)

NIETZSCHE

~~GOTT~~ IST TOT!

~~NIETZSCHE~~
GOTT

BEISPIEL CMI

Eine Mini-Ontologie für Briefkorrespondenzen nach TEI/CMI

CMI Ausgangsbasis:

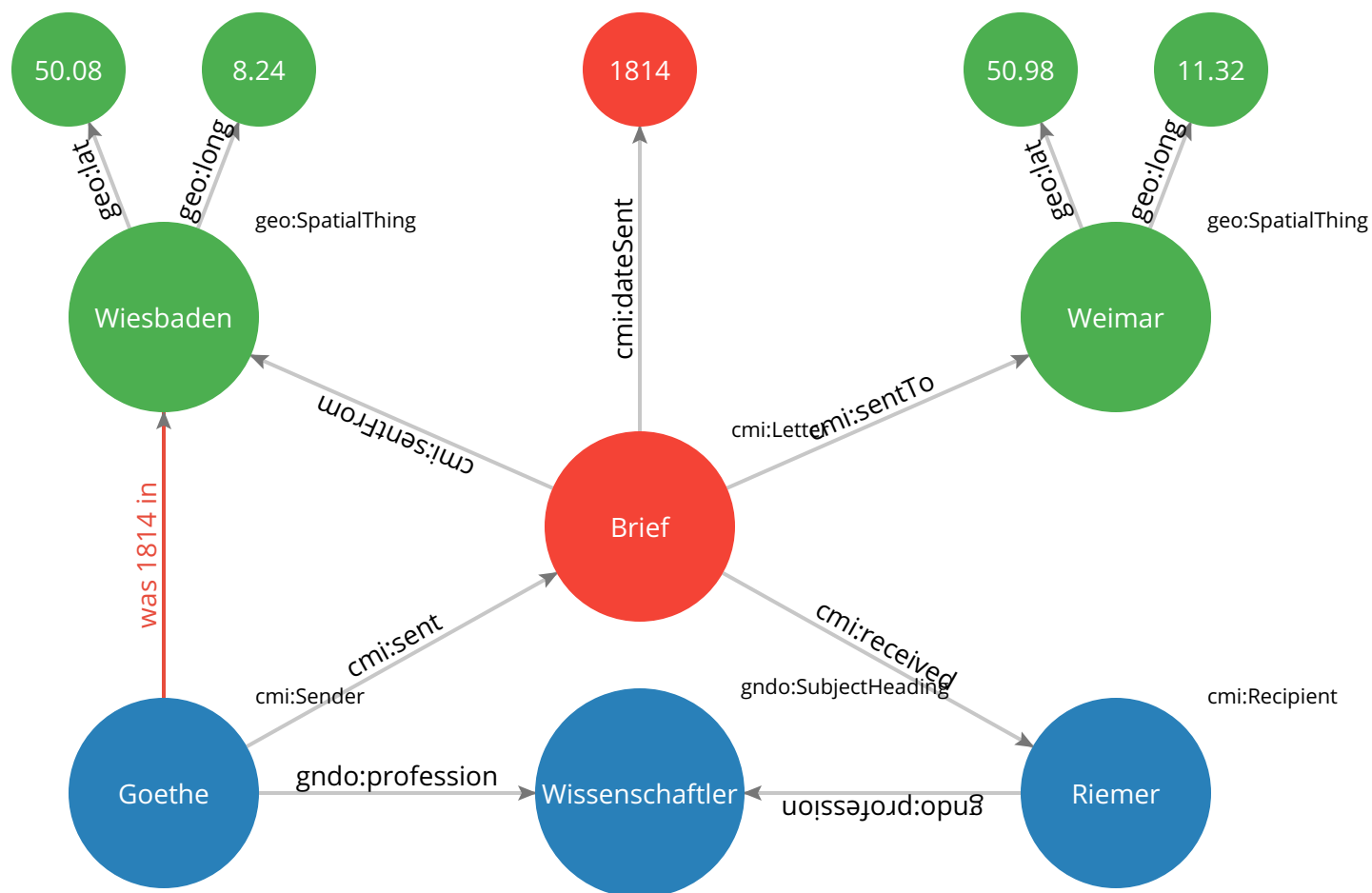
```
<correspDesc cs:source="#WEGA" ref="http://www.weber-gesamtausgabe.de/A040654">
  <correspAction type="sent">
    <persName ref="http://d-nb.info/gnd/118540238">Goethe, Johann Wolfgang von</persName>
    <placeName ref="http://www.geonames.org/2809346">Wiesbaden</placeName>
    <date when="1814-08-29"/>
  </correspAction>
  <correspAction type="received">
    <persName ref="http://d-nb.info/gnd/116543671">Riemer, Friedrich Wilhelm</persName>
    <placeName ref="http://www.geonames.org/2812482">Weimar</placeName>
  </correspAction>
</correspDesc>
```

Quelle in correspSearch

ANWENDUNG

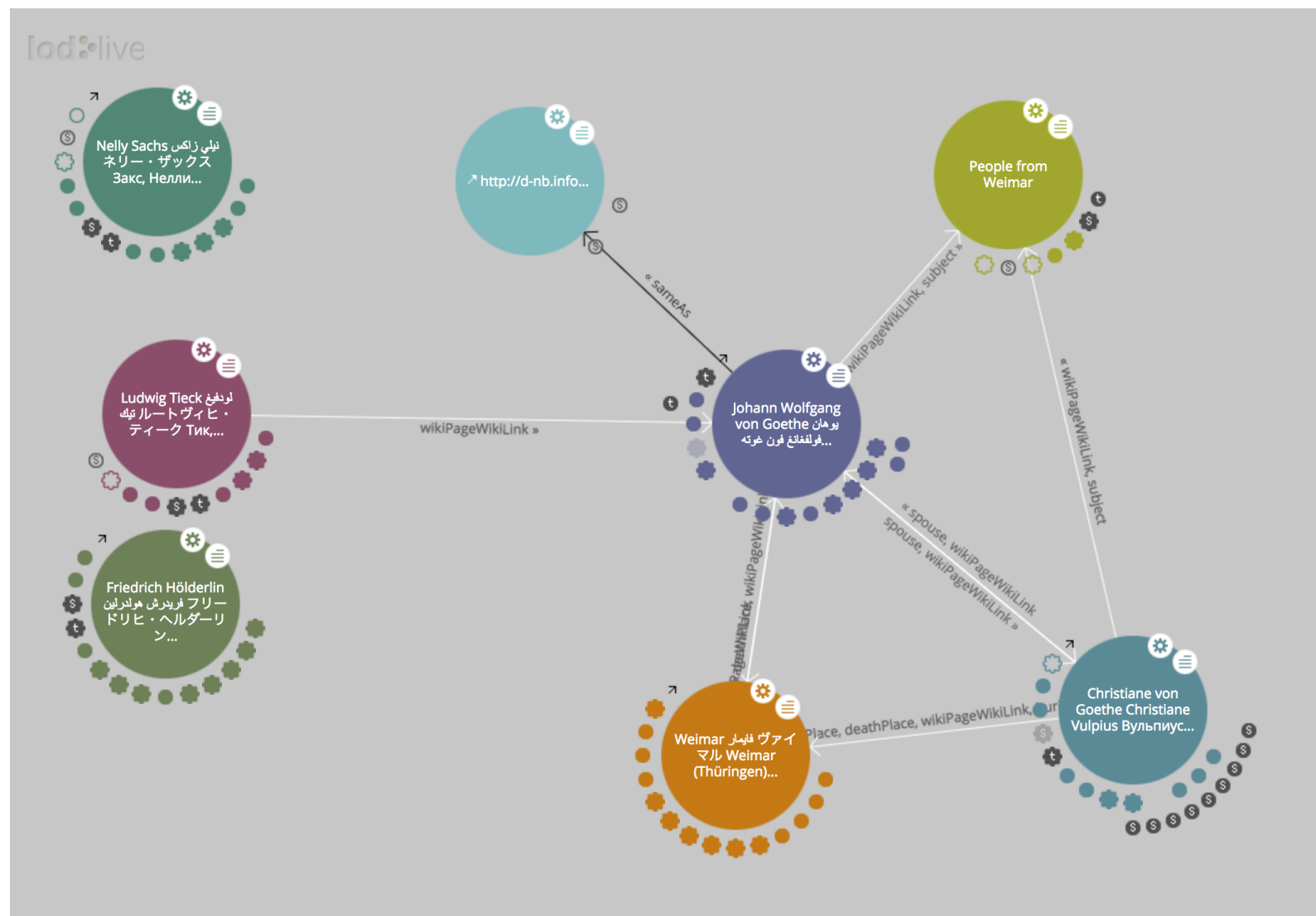
Goethe an Riemer

■ correspSearch ■ GND ■ Geonames



LOD LIVE

Explorationsmöglichkeit für den „Giant Global Graph“

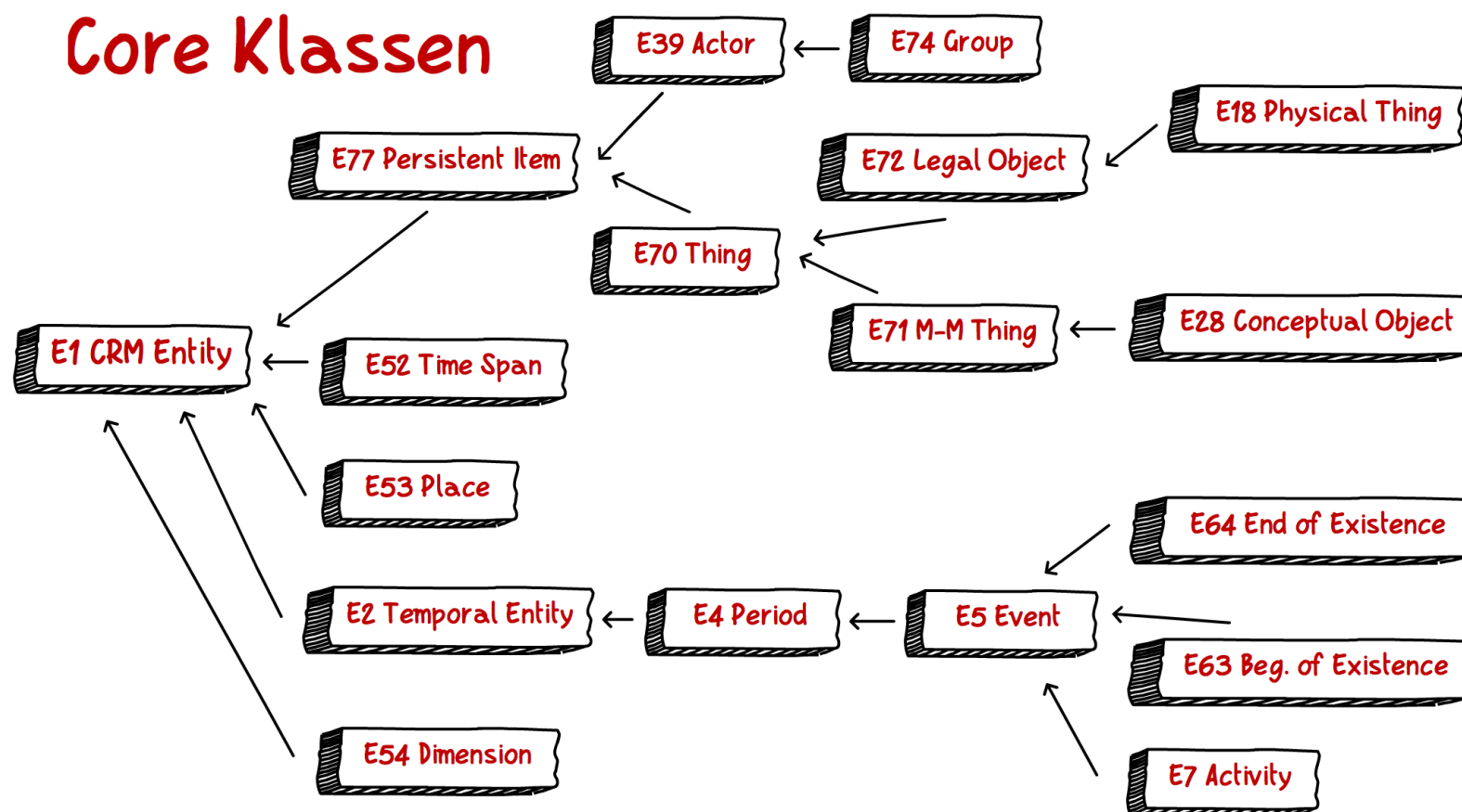


Quelle: <http://en.lodlive.it/>

CIDOC-CRM

Eine Maxi-Ontologie für Objekte des kulturellen Erbes

CIDOC-CRM Core Klassen



Erlangen-CRM Dokumentation

03 Praxis

ANWENDUNGSBEISPIELE

EPIGRAPHISCHE FACHDATEN

Epidat - Forschungsplattform für jüdische Grabsteinepigraphik

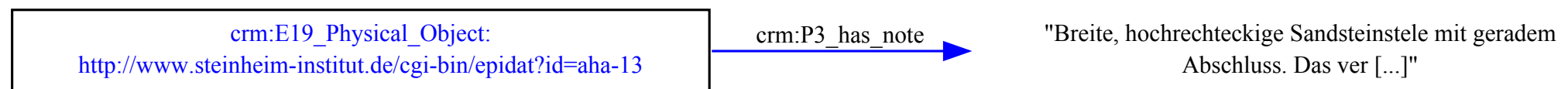


- ▶ 182 Sammlungen mit 33.279 Grabmalen und 65.074 Bilddigitalisaten (CC-BY 4.0)
- ▶ Maschinenlesbare Ausgabe der Datensätze als **EpiDoc-XML**
- ▶ **BEACON** Datei für im Bestand vorkommende Personen
- ▶ **HTTP-Schnittstelle** zum Harvesting der Datensätze (**Beispiel**)
- ▶ <http://www.steinheim-institut.de/cgi-bin/epidat>

SCHRITT 1

Grabstein 13 aus dem Epidat Corpus

Link: [Grabstein des Elieser und Jehuda ben Josef Hakohen \(1879\) | Erlangen CRM](http://www.steinheim-institut.de/cgi-bin/epidat?id=aha-13)

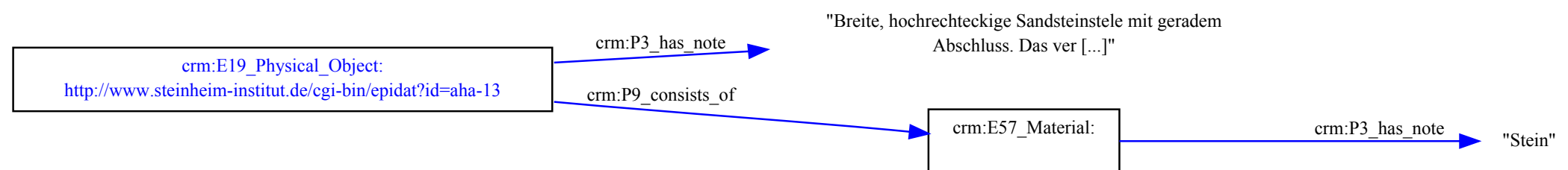


Daten: [SVG](#) | [RDF/XML](#) | [Turtle](#)

SCHRITT 2

Grabstein 13 aus dem Epidat Corpus

Link: [Grabstein des Elieser und Jehuda ben Josef Hakohen \(1879\) | Erlangen CRM](http://www.steinheim-institut.de/cgi-bin/epidat?id=aha-13)

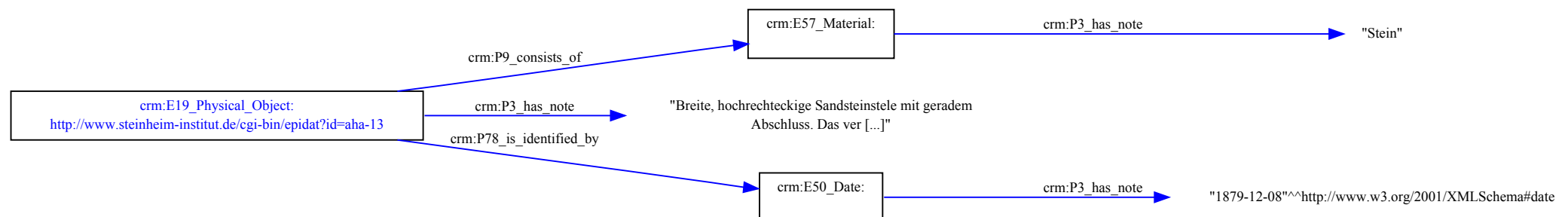


Daten: [SVG](#) | [RDF/XML](#) | [Turtle](#)

SCHRITT 3

Grabstein 13 aus dem Epidat Corpus

Link: [Grabstein des Elieser und Jehuda ben Josef Hakohen \(1879\) | Erlangen CRM](#)

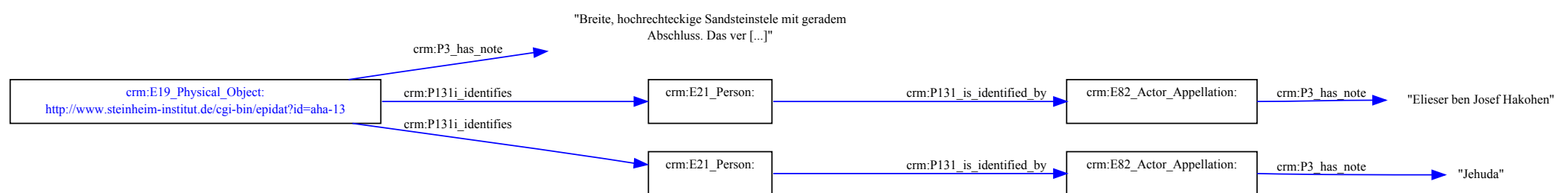


Daten: [SVG](#) | [RDF/XML](#) | [Turtle](#)

SCHRITT 4

Grabstein 13 aus dem Epidat Corpus

Link: [Grabstein des Elieser und Jehuda ben Josef Hakohen \(1879\)](http://www.steinheim-institut.de/cgi-bin/epidat?id=aha-13) | [Erlangen CRM](#)

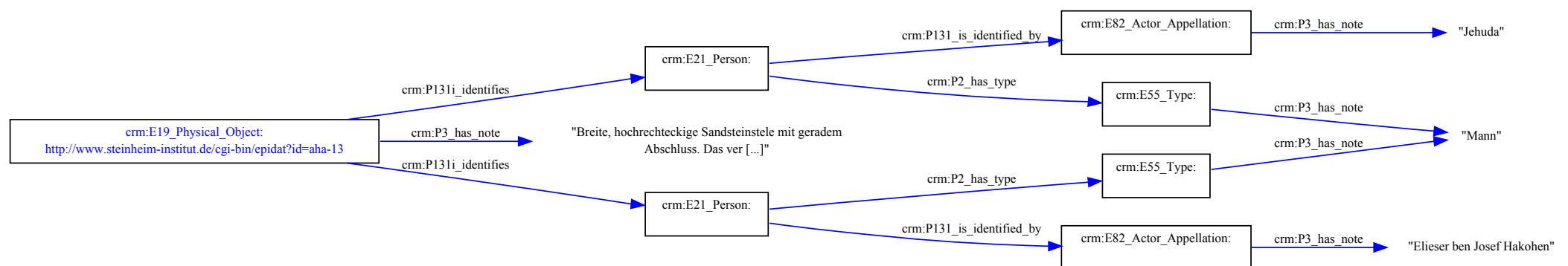


Daten: [SVG](#) | [RDF/XML](#) | [Turtle](#)

SCHRITT 5

Grabstein 13 aus dem Epidat Corpus

Link: Grabstein des Elieser und Jehuda ben Josef Hakohen (1879) | Erlangen CRM

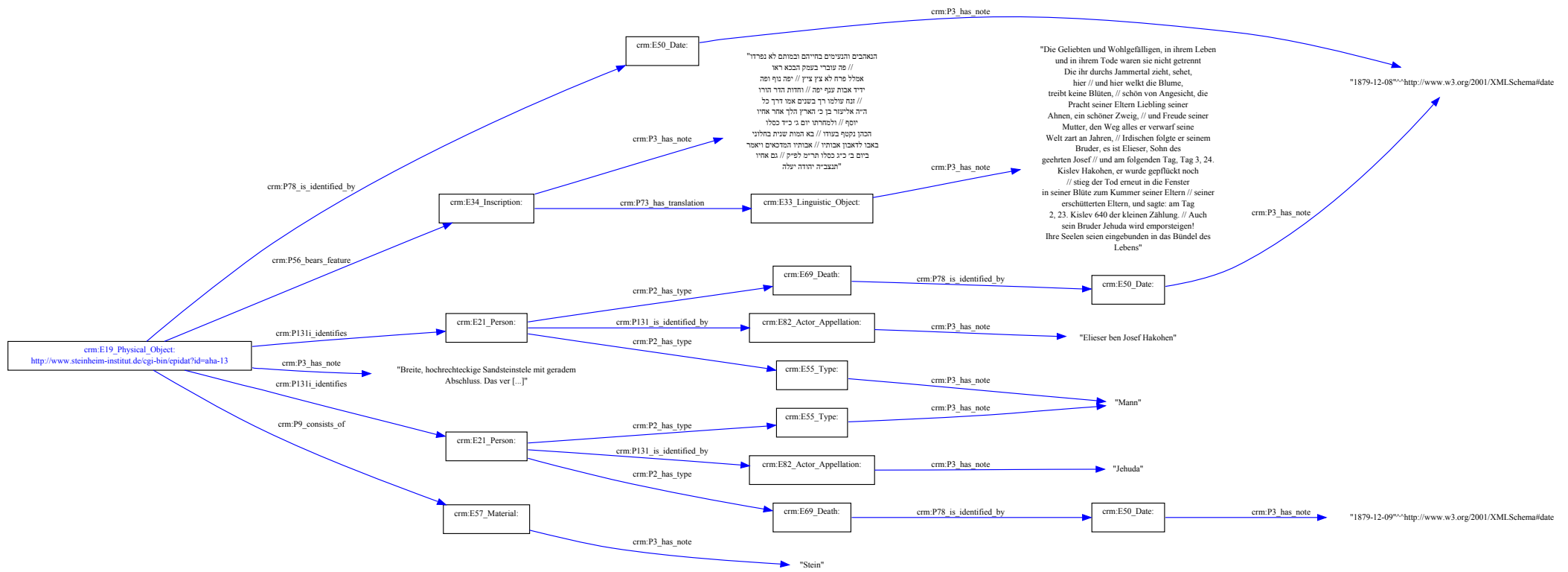


Daten: SVG | RDF/XML | Turtle

SCHRITT 6

Grabstein 13 aus dem Epidat Corpus

Link: [Grabstein des Elieser und Jehuda ben Josef Hakohen \(1879\) | Erlangen CRM](#)



Daten: SVG | RDF/XML | Turtle

GESAMTERGEBNIS

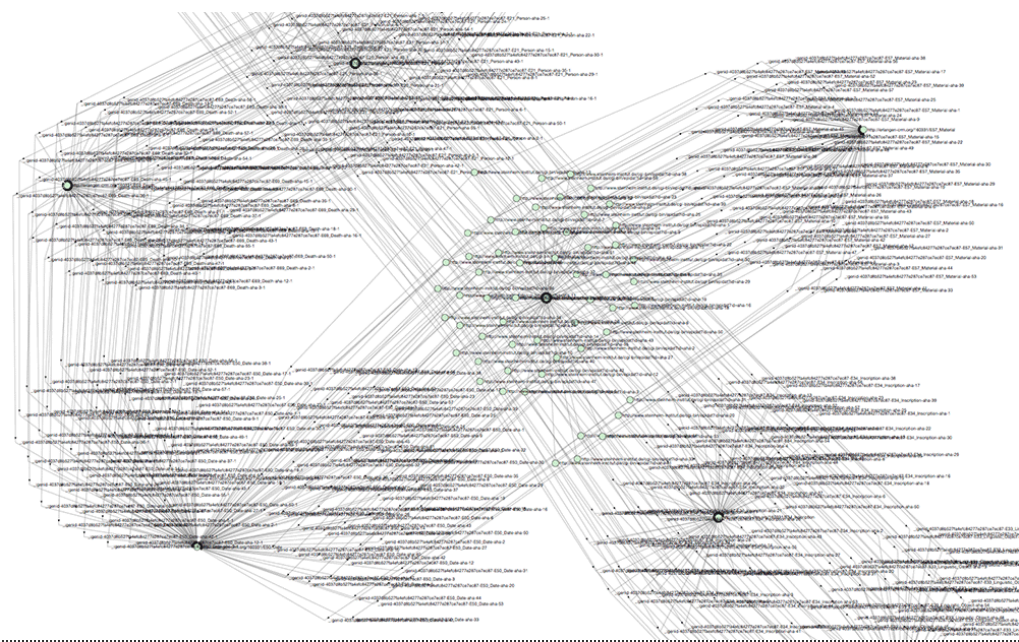
Grabsteinkorpus Ahaus

Statistik

Triple insgesamt: 1536

Klassen	Instanzen
E19_Physical_Object	57
E57_Material	1 x Kunststele, 3 x Stein, 21 x Muschelkalk, 25 x Sandstein, 7 x Kunststein
E82_Actor_Appellation	57
E55_Type	30 Männer / 27 Frauen
E69_Death	57

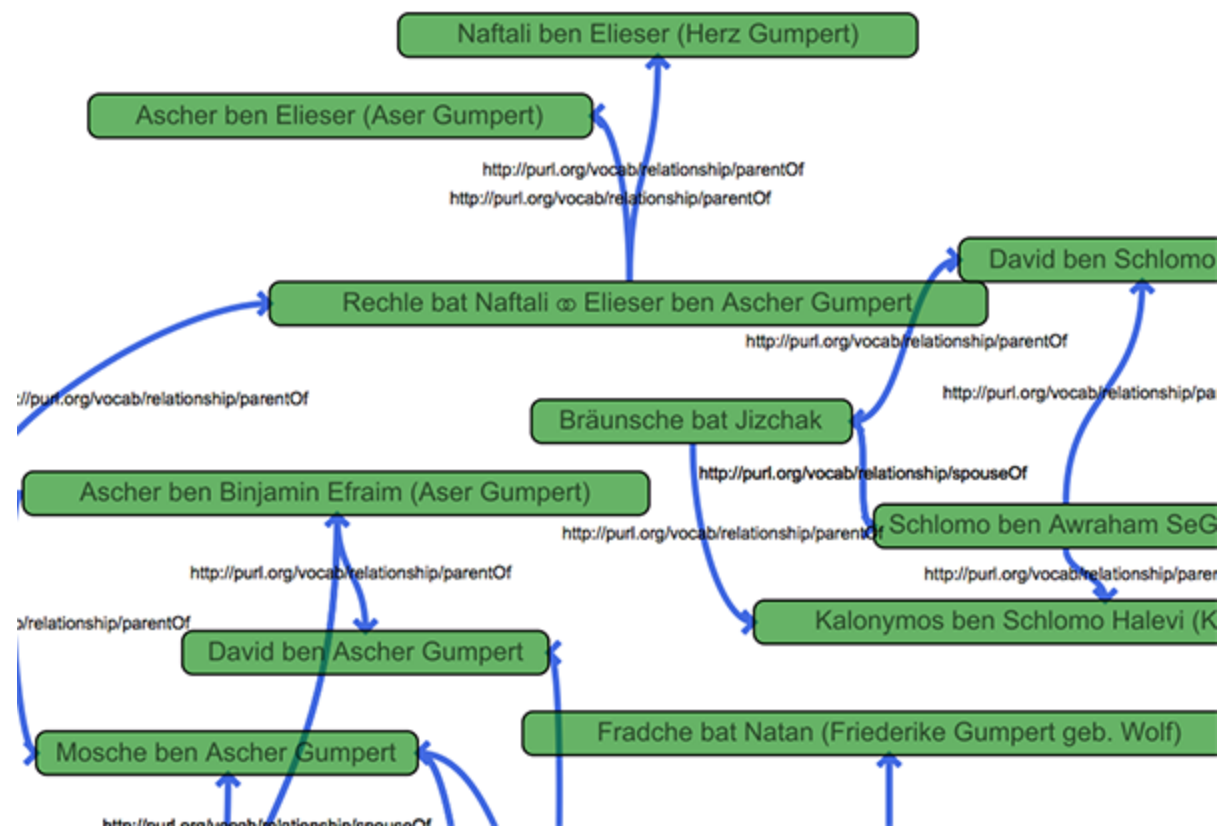
Vollständiger Graph



Daten aufrufen: [RDF/XML](#) | [SVG](#) | [Graph](#)

VISUALISIERUNG 1

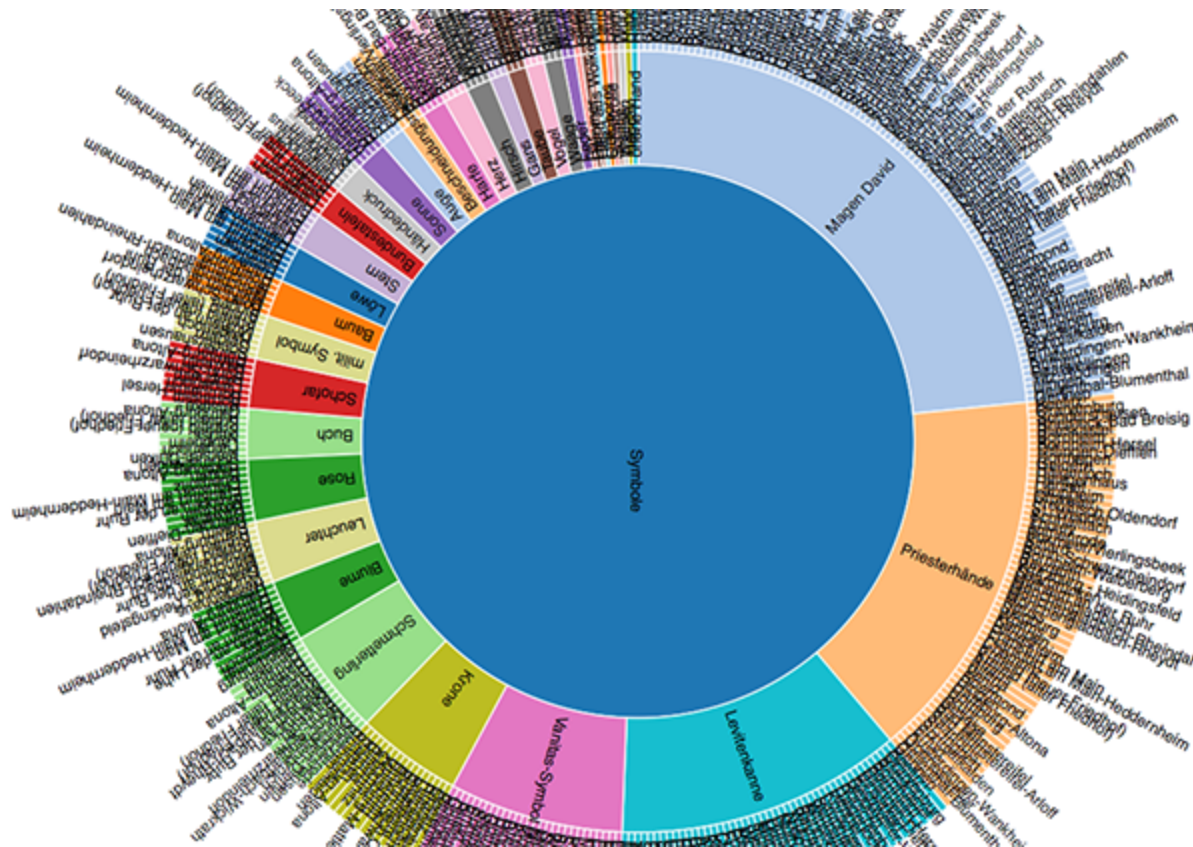
Epidat: Familienbeziehungen auf dem Friedhof Ahaus



Graph-Visualisierung aufrufen

VISUALISIERUNG 2

Epidat: Jüdische Grabsteinsymbole pro Friedhof



Sunburst-Visualisierung aufrufen

GUTENBERG BIOGRAPHICS

Bereitstellung strukturierter Daten durch RDFa Einbettung in HTML

```
<div typeof="Person" vocab="http://schema.org/" class="row">
  <h1><span class="given-name" property="givenName">Karl </span>
    <span class="family-name" property="familyName">Bechert</span>
  </h1>
  <p>Geb. <time datetime="1901-08-23" property="birthDate">23.08.1901</time>
    in <span typeof="Place" property="birthPlace">Nürnberg</span>
  </p>
  <p>Gest. <time datetime="1981-04-01" property="deathDate">01.04.1981</time>
    in <span typeof="Place" property="deathPlace">Weilmünster-Möttau</span>
  </p>
  <p><span class="gnd">GND:
    <a href="http://d-nb.info/gnd/119127059" property="sameAs">119127059</a>;
    VIAF: <a href="http://viaf.org/viaf/32323656" property="sameAs">32323656</a></span>
  </p>
  <ul>
    <li property="performerIn" typeof="Event">
      <span property="name">
        <time property="startDate" datetime="1946-01-01" class="skip">1946-01-01</time>
        <time property="endDate" datetime="1969-12-31" class="skip">1969-12-31</time>
        1946-1969, Professor für Theoretische Physik, Naturwissenschaftliche Fakultät
      </span>
    </li>
  </ul>
  [...]
</div>
```

Beispiel [Karl Bechert](#) aus dem [Mainzer Professorenkatalog](#) | Google Test-Tool: [Demo](#)

- ▶ Einfach umzusetzende maschinenlesbare Datenanreicherung
- ▶ Daten können direkt aus den HTML-Frontends abgegriffen und umgewandelt (bspw. in JSON-LD) werden
- ▶ Steigert gleichzeitig die Suchmaschinenoptimierung der Kataloge

DEUTSCHE INSCRIFTEN ONLINE

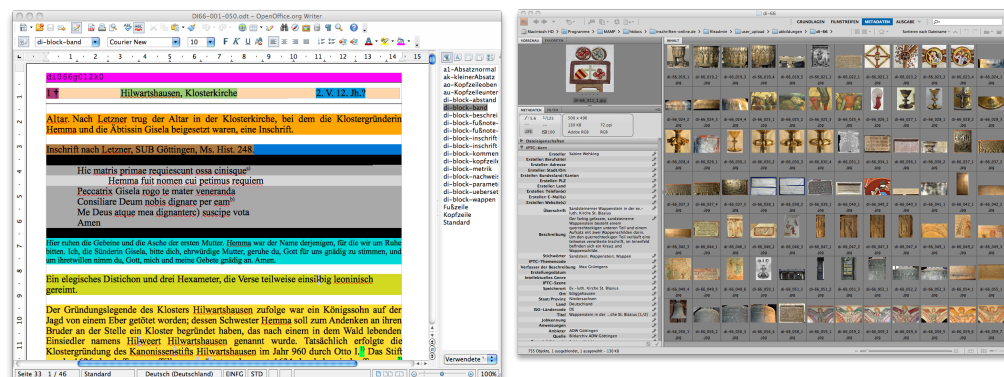
Einige Funktionalitäten im Überblick



- ▶ 42 Inschriftenbände bzw. -bestände mit ca. 17.000 Katalognummern und 18.000 Abbildungen
- ▶ Flexible XML Workflows zur Überführung der Banddaten in die Applikation
- ▶ Persistente Referenzierung aller Artikel mittels URN der Deutschen Nationalbibliothek
- ▶ Standardkonforme Ausgabe der Kataloartikel als TEI/EpiDoc via REST-

Interface

<http://www.inschriften.net/>

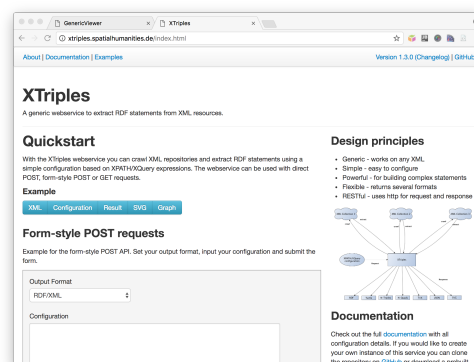
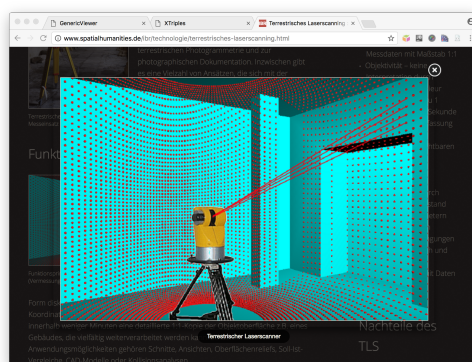


BMBF PROJEKT (2012-2015)

Inschriften im Bezugssystem des Raumes



- ▶ Verbindung von Raumdaten (Terrestrisches Laserscanning) mit Fachdaten (RDF) mit dem Ziel neuer Analysemöglichkeiten
- ▶ Modulare Architektur, gekoppelt über REST-basierte Schnittstellen
- ▶ Webbasiertes Werkzeug zur Erfassung und semantischen Verknüpfung von Raumgeometrien in der Punktwolke (Generic Viewer)

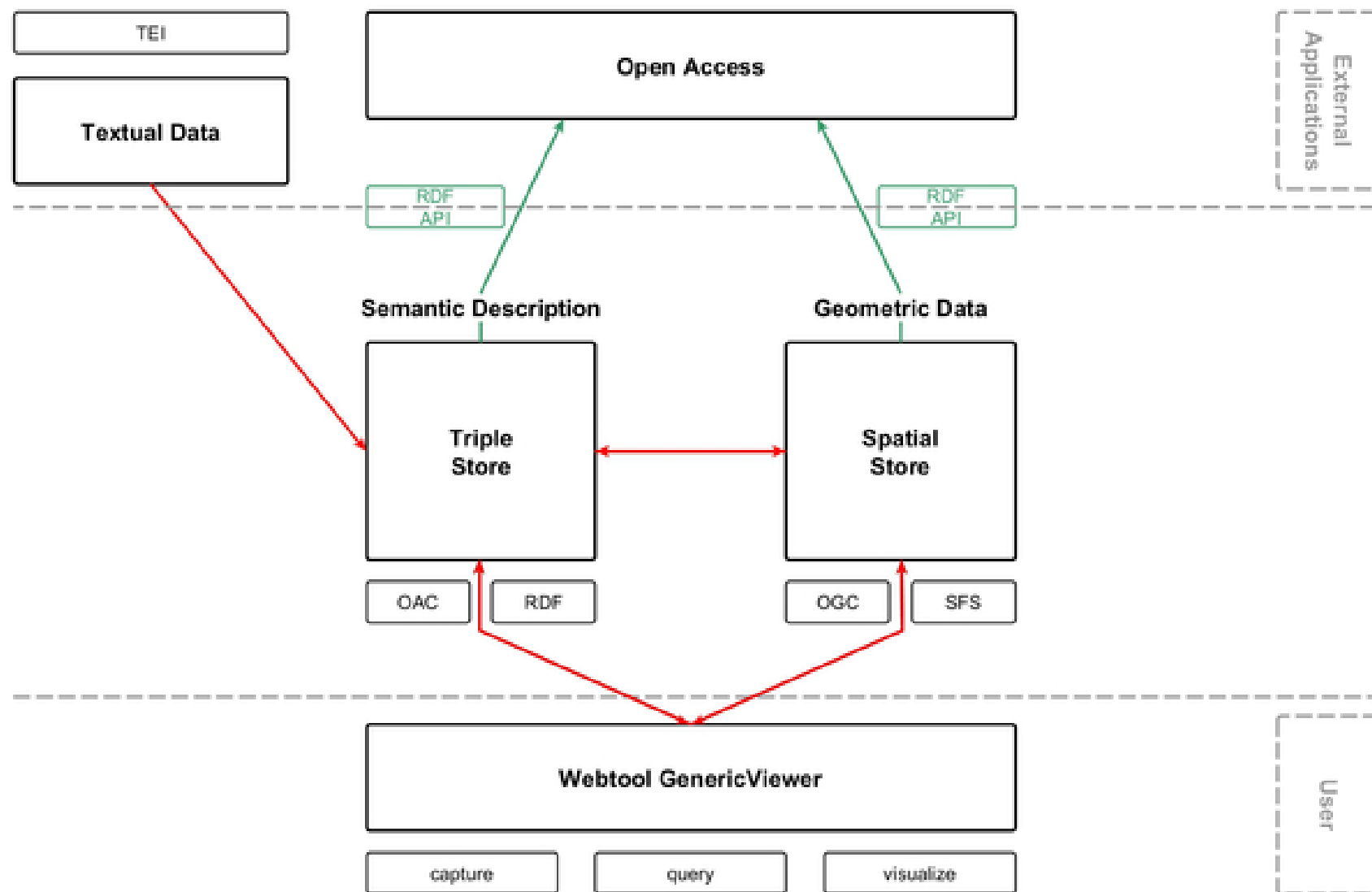


- ▶ Webservice zum Harvesting und zur Extraktion semantischer Statements aus XML-basierten Fachdaten (XTriples)

<http://spatialhumanities.de/>

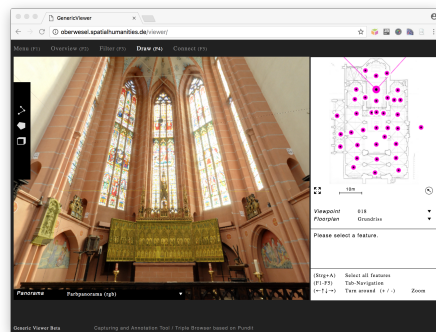
IBR ARCHITEKTUR

Komponenten

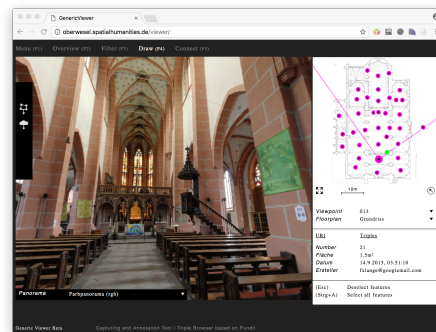


GENERIC VIEWER

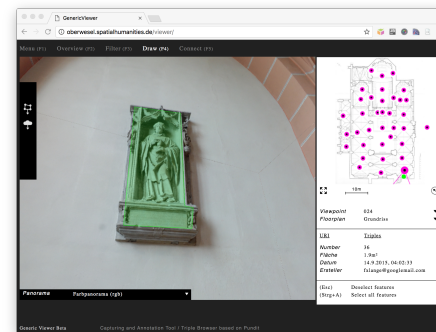
Funktionen



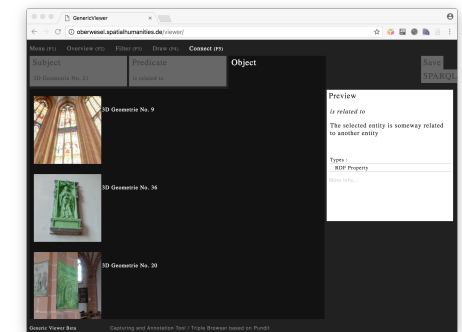
Orientierung



Messung



Erfassung



Verknüpfung

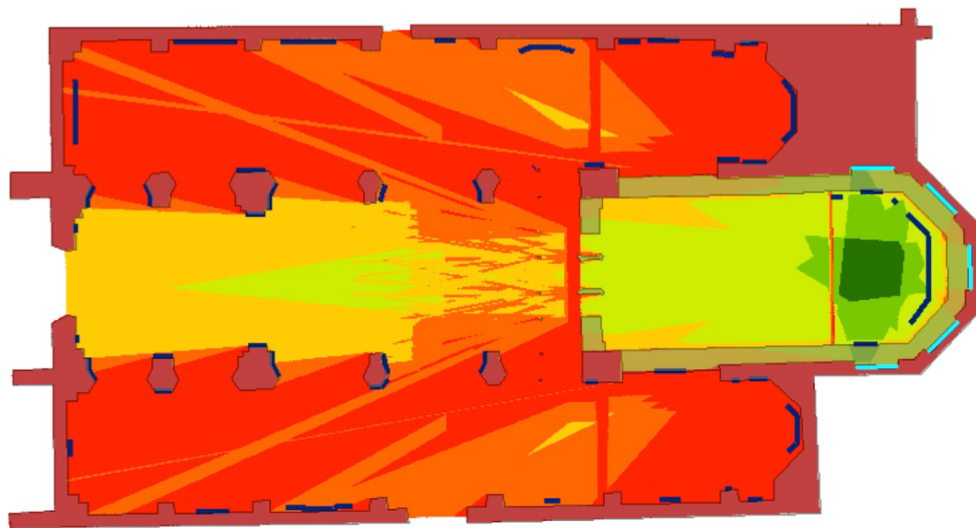
Ausprobieren:

<http://oberwesel.spatialhumanities.de/viewer/>

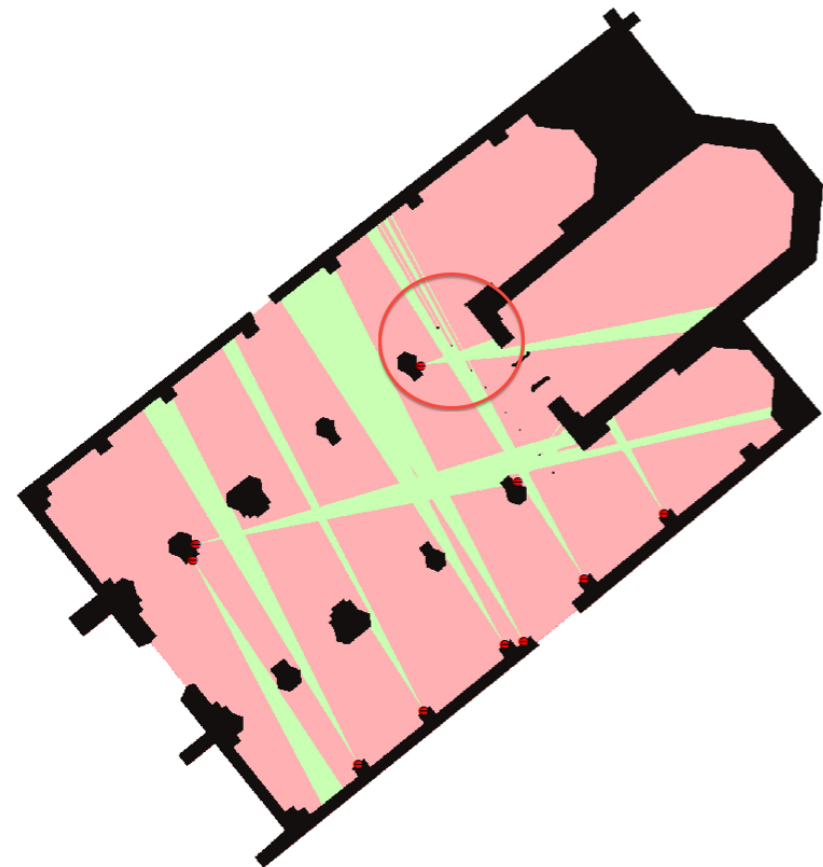
REFERENZSTUDIEN

Liebfrauenkirche Oberwesel

Sichtbarkeit der Bauinschrift



Petrus Lutern / Martha-Altar



Paper: Lange, Unold (2015): Semantisch angereicherte 3D-Messdaten von Kirchenräumen als Quellen für die geschichtswissenschaftliche Forschung, ZFDG 1 (2015), http://www.zfdg.de/sb001_015

04 Zusammenfassung

POTENTIALE VON LOD IN DEN DIGITAL
HUMANITIES

ANALYSE

Einige Fragestellungen, bei deren Beantwortung LOD helfen kann

- ▶ Welche historischen Personen existieren repositorienübergreifend in welchen spezifischen Kontexten?
- ▶ Welche kulturellen Objekte existieren zu welchen Zeiten in den verschiedenen Fachdatenkorpora?
- ▶ Können bestimmte Muster im Auftreten bestimmter historisch-kultureller Phänomene aus den Daten gefolgert werden?
- ▶ Auf Ebene der Metadaten: Welche geisteswissenschaftlichen Ressourcen beinhaltet ein Repository, wie sind diese strukturiert und in welcher Relation stehen sie zueinander?

VERNETZUNG UND NACHNUTZUNG

LOD für Fachdaten

Fachdaten und/oder Metadaten zu diesen Fachdaten als LOD bereitzustellen ermöglicht

- ▶ ... es Dritten, meine digitalen Ressourcen zu finden.
- ▶ ... einen freien, strukturierten und maschinenlesbaren Zugriff (ggf. zu eigenen Analyse Zwecken).
- ▶ ... eine standardisierte Bereitstellung der Informationen nach W3C- und ISO-Standards.
- ▶ ... die Daten in einen größeren Kontext zu stellen und Datensilos somit aufzulösen.
- ▶ **Und zwar ohne Kontrollverlust über die eigenen Daten und deren Qualität!**

FINIS

DANKE FÜR DIE AUFMERKSAMKEIT

Software & Attribution

LINKS, SOFTWARE & ATTRIBUTION

Software

- ▶ [Impress.js](#) (Presentation)
- ▶ [Open Sans](#) (Google Font)
- ▶ [Skeleton CSS](#) (CSS Micro Framework)
- ▶ [Magnific Popup](#) (Lightbox)
- ▶ [greuler](#) (Graph Modelling)
- ▶ [highlight.js](#) (Syntax Highlighting)
- ▶ [decktape](#) (PDF Export)

Attribution

- ▶ Lizenz: [CC-BY 4.0](#), Torsten Schrade, mit Slides, Anregungen und Ideen von [Andreas Wagner](#) und [Thomas Kollatz](#)